Insights from the Genome Sequence of a *Salmonella enterica* Serovar Typhi Strain Associated with a Sporadic Case of Typhoid Fever in Malaysia

Kien-Pong Yap, Cindy Shuan Ju Teh, Ramani Baddam, Lay-Ching Chai, Narender Kumar, Tiruvayipati Suma Avasthi, Niayz Ahmed and Kwai-Lin Thong

Institute of Biological Sciences, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia; Laboratory of Biomedical Science and Molecular Microbiology, UMBIO Research Cluster, University of Malaya, Kuala Lumpur, Malaysia; and Pathogen Biology Laboratory, Department of Biotechnology, School of Life Sciences, University of Hyderabad, Hyderabad, India

*Salmonella enterica* serovar Typhi is the causative agent of typhoid fever, which causes nearly 21.7 million illnesses and 217,000 deaths globally. Herein, we describe the whole-genome sequence of the *Salmonella* Typhi strain ST0208, isolated from a sporadic case of typhoid fever in Kuala Lumpur, Malaysia. The whole-genome sequence and comparative genomics allow an in-depth understanding of the genetic diversity, and its link to pathogenicity and evolutionary dynamics, of this highly clonal pathogen that is endemic to Malaysia.

*Salmonella enterica* serovar Typhi (S. Typhi) is a human intra-cellular pathogen of global importance, infecting 21.7 million people and causing 217,000 deaths annually (4). The challenges of higher incidence of typhoid fever in developing countries have led to an increased health burden (4).

In 2008, 201 cases of sporadic outbreaks were reported in Malaysia (18). The genome sequence of this sporadically associated strain will provide insights into possible genetic events that would confer a fitness advantage.

S. Typhi strain ST0208 was isolated from the stool sample of a typhoid fever patient admitted to University Malaya Medical Centre (UMMC), Kuala Lumpur, Malaysia, in 2008. The strain was characterized by pulsed-field gel electrophoresis (PFGE), repetitive extragenic palindromic (REP)-PCR, and antimicrobial susceptibility profiling (25). The genome sequence of *S. Typhi* ST0208 was determined using an Illumina genome analyzer (GA2x, pipeline version 1.60) with an insert size of 300 bp, which generated 1.83 gigabytes of data with an average coverage of 165× and yielded 1,499,986 paired-end reads with a 100-bp read length. Genome assembly was constructed de novo using Velvet (26), which generated 222 contigs. The resultant contigs were uploaded into the RAST server (2, 17) to predict the open reading frames (ORFs) by using Glimmer3 (5) and validated with the ISGA integrated system (8). In brief, the predicted ORFs were annotated by searching against clusters of orthologous group (21) and SEED (7) databases, whereas tRNA and rRNA genes were identified by using tRNAscAn-SE (15) and RNaMMer (12), respectively. The draft genome size is approximately 4,798,272 bp in length, with an average GC content of 52.0%, and is composed of 4,890 predicted coding sequences with an average length of 810 bp. A mean percentage of 83.7% of nucleotides of the genome are predicted to encode proteins. The genome reveals 71 tRNA and 22 rRNA predicted genes.

The genome contains several monosaccharide and polysaccharide metabolism-related genes, which were not reported in *S. Typhi* strains Ty2 and CT18 (6, 11, 18), such as D-galactarate permease, gluconate permease, tagatose-6-phosphate kinase, trehalase, and arabinose-proton transporter, that could be associated with host persistence (20). The genome sequence revealed four multidrug resistance clusters, *mdtABCR* and *marABCR* proteins (1, 3, 19), DNA gyrase subunit A and B, and topoisomerase subunit (IV) A and B (9), which were also identified in *S. Typhi* Ty2 and *S. Typhi* CT18 (6, 11, 19). Hypothetical proteins and pathogenicity islands were annotated.

High genetic diversity of *S. Typhi* was detected among human strains in Malaysia and Southeast Asia (10, 14, 22, 23, 24). Variation in the genome sequence revealed by the strain ST0208 is consistent with the proposed key theory of persistent adaptation and optimization of function (13, 16, 19). Genomic information from locality-specific strains associated with clinical manifestation allows genome evolution and endemicity to be studied.

**Nucleotide sequence accession numbers.** This whole-genome shotgun project has been deposited in DDBJ/EMBL/GenBank under the accession number AJXA00000000. The version described in this paper is the first version, AJXA01000000. The BioProject designation for this project is PRJNA160181.

**ACKNOWLEDGMENTS**

This research is supported by the University of Malaya High Impact Research Grant (reference UMC/1625/1HIR/MOHE/-02 [A000002-5000 1]), Molecular Genetics, in which K.-L. Thong is the grant holder for the high-impact research project under the title of Pathogenomics and Phenomics of Food-borne Bacterial Diseases.

We are also grateful to M’s Genotypic Technology Pvt. Ltd., Bengaluru, India, for their assistance with Illumina sequencing. We acknowledge Safwan Jusoh from the ICT department, University of Malaya, for assisting us with computing solutions and allowing us to use the servers and computational facilities and S. T. Ngoi from LBSSM, IPS, for technical assistance with DNA preparation.

**REFERENCES**


